Open access • Posted Content • DOI:10.1101/2020.07.17.209023

Blood-derived mitochondrial DNA copy number is associated with gene expression across multiple tissues and is predictive for incident neurodegenerative disease

Stephanie Y. Yang, Christina A. Castellani, Ryan J. Longchamps, Vamsee Pillalamarri ...+3 more authors

Institutions: Johns Hopkins University School of Medicine, Johns Hopkins University

Published on: 18 Jul 2020 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Mitochondrial DNA replication, Gene expression, Gene, NRF1 and Whole blood

Blood-derived mitochondrial DNA copy number is associated with gene expression across multiple tissues and is predictive for incident neurodegenerative disease

Stephanie Y. Yang[1], Christina A. Castellani[1], Ryan J. Longchamps[1], Vamsee K. Pillalamarri[1], Brian O'Rourke[2], Eliseo Guallar[3], Dan E. Arking[1,2]

[1]McKusick-Nathans Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD
[2]Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD
[3]Departments of Epidemiology and Medicine, and Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD

Email addresses:
syang93@jhmi.edu
ccastell8@jhmi.edu
rlongcha@broadinstitute.org
vpillal1@jhmi.edu
bor@jhmi.edu
eguallar@jhmi.edu
arking@jhmi.edu

*Correspondence and address for reprints to:
Dan E. Arking, Ph.D.
Johns Hopkins University School of Medicine
733 N Broadway Ave
Miller Research Building Room 459
Baltimore, MD 21205
(410) 502-4867 (ph)
arking@jhmi.edu

ABSTRACT

Background

Mitochondrial DNA copy number (mtDNA-CN) can be used as a proxy for mitochondrial function and is associated with a number of aging-related diseases. However, it is unclear how mtDNA-CN measured in blood can reflect risk for diseases that primarily manifest in other tissues. Using the Genotype-Tissue Expression Project, we interrogated the relationships between mtDNA-CN measured in whole blood and gene expression from whole blood as well as 47 additional tissues.

Results

We evaluated associations between blood-derived mtDNA-CN and gene expression in whole blood for 418 individuals, correcting for known confounders and surrogate variables derived from RNA-sequencing. Using a permutation-derived cutoff ($p<2.70e-6$), mtDNA-CN was significantly associated with expression for 721 genes in whole blood, including nuclear genes that are required for mitochondrial DNA replication. Significantly enriched pathways included splicing ($p=1.03e-8$) and ubiquitin-mediated proteolysis ($p=2.4e-10$). Genes with target sequences for the mitochondrial transcription factor NRF1 were also enriched ($p=1.76e-35$).

In non-blood tissues, there were more significantly associated genes than expected in 30 out of 47 tested tissues, suggesting that global gene expression in those tissues is correlated with mtDNA-CN. Pathways that were associated in multiple tissues included RNA-binding, catalysis, and neurodegenerative disease. We evaluated the association between mtDNA-CN and incident neurodegenerative disease in an independent dataset, the UK Biobank, using a Cox proportional-hazards model. Higher mtDNA-CN was significantly associated with lower risk for incident neurodegenerative disease (HR=0.73, 95% CI= 0.66;0.90).

Conclusions

The observation that mtDNA-CN measured in whole blood is associated with gene expression in other tissues suggests that blood-derived mtDNA-CN can reflect metabolic health across multiple tissues. Key pathways in maintaining cellular homeostasis, including splicing, RNA binding, and catalytic genes were significantly associated with mtDNA-CN, reinforcing the importance of mitochondria in aging-related disease. As a specific example, genes involved in neurodegenerative disease were significantly enriched in multiple tissues. This finding, validated in a large independent cohort study showing an inverse association between mtDNA-CN and neurodegenerative disease, solidifies the link between blood-derived mtDNA-CN, altered gene expression in both blood and non-blood tissues, and aging-related disease.

BACKGROUND

Mitochondria perform multiple essential metabolic functions including energy production, lipid metabolism, and signaling for apoptosis. Mitochondria possess circular genomes (mtDNA) that are distinct from the nuclear genome. While cells typically only possess two copies of the nuclear genome, they contain 100s to 1000s of mitochondria, and each individual mitochondrion can hold 2-10 copies of mtDNA resulting in wide variation in mtDNA copy number (mtDNA-CN) [1]. The amount of mtDNA-CN also varies widely across cell types, with higher energy demand cell types typically possessing higher levels of mtDNA-CN [1–3]. Due to the importance of mitochondria in metabolism and energy production, mitochondrial dysfunction plays a role in the etiology of many human diseases [4]. mtDNA-CN has been shown to be a proxy for mitochondrial function, and is consequently an attractive biomarker due to its ease of measurement [5,6]. Indeed, low levels of mtDNA-CN in peripheral blood have been associated with an increased risk for a number of chronic aging-related diseases including frailty, kidney disease, cardiovascular disease, heart failure, and overall mortality [7–10].

Crosstalk between the mitochondrial and nuclear genomes is essential for maintaining cellular homeostasis. Many essential mitochondrial proteins are encoded by the nuclear genome, and expression of these nuclear genes must be modified to match mitochondrial activity. Likewise, mitochondrial activity must respond to cellular energy demands. Polymorphisms in the nuclear genome have been associated with changes in mitochondrial gene expression, and mitochondrial genome variation has been associated with changes in nuclear gene expression, suggesting interplay between the two genomes [11,12].

In cancer cells, mtDNA-CN alters gene expression through modifying DNA methylation [13,14]. Recent work from our lab has shown that mtDNA-CN is also associated with nuclear DNA methylation in noncancer settings [15]. Given that DNA methylation can modify gene expression, the current study seeks to explore the potential association between blood-derived mtDNA-CN and gene expression. Past

work has shown that mtDNA-CN is associated with gene expression of nuclear-encoded genes in lymphoblast cell lines, but this may not reflect biological processes occurring in other tissues, especially after an extended culturing period [16]. Therefore, we leveraged data from the Genotype-Tissue Expression Project (GTEx), a cross-sectional study with gene expression data from multiple non-diseased postmortem tissues, to examine associations between mtDNA-CN and expression of both nuclear and mitochondrially-encoded genes [17]. We found that blood-derived mtDNA-CN was globally associated with increased gene expression in whole blood. Additionally, blood-derived mtDNA-CN was associated with gene expression in other, non-blood tissues across the body. Specifically, genes annotated with neurodegenerative disease pathways were significantly enriched, leading to a follow-up analysis that uncovered a novel association between blood-derived mtDNA-CN and incident neurodegenerative disease.

RESULTS

Determination and validation of mtDNA-CN metric

mtDNA-CN estimates were generated from whole genome sequences performed on DNA derived from whole blood using the ratio of mitochondrial reads to total aligned reads. As mtDNA-CN is known to be affected by cell type composition, cell counts for samples with available RNA-sequencing data were deconvoluted using gene expression measured in whole blood [18,19]. We identified a batch effect that resulted in significantly altered mtDNA-CN for individuals sequenced prior to January 2013. Therefore, only individuals sequenced after January 2013 were retained for analysis (Supp. Fig. 1). After quality control, outlier filtering, and normalization of the RNA-sequencing data, 418 individuals remained for analyses (see **Methods**).

To validate mtDNA-CN measurements in the filtered GTEx data, we determined the association between mtDNA-CN and known correlated measures, including age, sex, and neutrophil count [18,20,21]. We

observed a significant association with neutrophil count (p=5e-05), with higher neutrophil count associated with lower mtDNA-CN. While not statistically significant, effect size estimates between mtDNA-CN and age (p=0.19) and sex (p=0.17) were also in the expected direction. Effect sizes estimates for age and neutrophils were also consistent with prior literature [22] (Supp. Table 1). Based on variance explained from previous studies, the current study was only powered to detect a significant effect for neutrophil count. For all downstream analyses, mtDNA-CN was the standardized residual from a linear regression model adjusted for age, sex, cell counts, ischemic time, and cohort (see **Methods**).

Association of mtDNA-CN derived from whole blood with gene expression in blood

A priori, we expect that mitochondrially encoded gene expression would be positively correlated with mtDNA-CN. Likewise, multiple nuclear encoded genes are involved in the regulation of mtDNA replication, and thus, expression levels of these genes are expected to be correlated with mtDNA-CN [23,24]. We therefore evaluated the associations between mtDNA-CN and expression of these two classes of genes, correcting for cohort, sample ischemic time, genotyping PCs, age, race, and surrogate variables derived from RNA-sequencing data to capture known and hidden confounders (Supp. Fig. 2) [25].

To minimize the potential impact of outliers, we performed an inverse normal transformation on both the mtDNA-CN metric and the gene expression values. To evaluate the association between mtDNA-CN and mitochondrial RNA (mtRNA) levels, we used the median gene expression value calculated from scaled expression values across 36 mtDNA-encoded genes that passed expression thresholds (see **Methods**).

We observed a highly significant association between mtDNA-CN and overall mtRNA expression (p=9.10e-9) (Table 1), with 33 out of 36 individual mtDNA-encoded genes nominally significant (p<0.05) (Supp. Fig. 3).

Table 1. Blood-derived mtDNA-CN is positively associated with gene expression for mitochondrially encoded genes and nuclear encoded genes required for mtDNA replication.

Gene	Effect size estimate	Standard error	P-value
Scaled mtRNA median	0.15	0.03	9.10e-09
mtDNA replication machinery			
POLG	0.02	0.01	0.025
POLG2	0.06	0.02	4.08e-04
TWNK	0.03	0.02	0.11
SSBP1	0.06	0.01	1.38e-04
PRIMPOL	0.04	0.02	0.020
DNA2	0.05	0.02	0.010
MGME1	0.04	0.02	0.048
RNASEH1	0.06	0.02	2.51e-04
mtDNA transcription machinery			
TFAM	0.06	0.01	1.83e-04
TEFM	0.03	0.02	0.05
TFB2M	0.03	0.02	0.028
POLRMT	0.01	0.01	0.19
Nucleotide metabolism genes			
TK2	0.02	0.02	0.11
DGUOK	0.06	0.02	6.39e-04
RRM2B	0.04	0.01	0.006
TYMP	0.02	0.02	0.29
SLC25A4	0.08	0.03	0.003

† Genes from Rusecka et al.

Effect size estimates represent the change in gene expression, in standard deviation units, associated with a 1 standard deviation increase in blood-derived mtDNA-CN. Mitochondrially encoded genes are represented as the median of the scaled mtRNA expression of the 36 genes with detectable expression. Genes required for mtDNA replication were obtained from Rusecka et al [24].

In addition to genes coding directly for mtDNA replication machinery, genes involved in mtDNA transcription and nucleotide metabolism are also required for mtDNA replication. The mtDNA transcription machinery provides the RNA primers used in mtDNA replication and nucleotides are

needed to synthesize new mtDNA molecules. Of the 17 mtDNA major replication genes tested [24], all were positively associated with mtDNA-CN, as would be expected based on gene function; 8 of them were nominally significant ($p<0.05$), and were 4 significant after Bonferroni correction ($p<2.94e-3$) for multiple testing (Table 1).

To identify additional genes and pathways associated with mtDNA-CN, we performed a transcriptome-wide analysis. There was an overall inflation of test statistics, which we quantified using the genomic inflation factor (lambda = 4.71) [26]. Two-stage permutation testing demonstrated no inflation in null datasets, suggesting that this inflation represents a true global association between blood-derived mtDNA-CN and gene expression (Supp. Fig. 4).

When stratified by gene functional categories [27], all categories showed elevated test statistics, but protein-coding genes were the most enriched (lambda=7.44) (Fig. 1). Gene expression levels of most of the nominally significant genes were positively correlated with mtDNA-CN (7769 genes with positive t-values vs. 285 genes with negative t-values). While much of this positive skewing is due to correlated gene expression, permuted datasets demonstrate that this positive shift is significant ($P<0.001$, Supp. Fig. 5), perhaps reflecting a more active transcriptional state associated with higher mtDNA-CN. Only two negatively associated genes passed permutation cutoff ($p=2.7e-6$), *CAMP* ($p=1.58e-8$) and *PGLYRP1* ($p=1.78e-7$), both of which are involved in innate immunity.

Gene set enrichment analysis uncovers gene regulatory networks in whole blood

To identify specific molecular pathways, transcription factors, and gene ontologies associated with mtDNA-CN in whole blood, we performed gene set enrichment analyses [28] using gene sets obtained from the Molecular Signatures database (MSigDB) [29–33]. Previous studies have shown that cross-mappability can lead to false pseudogene positives in eQTL association studies [34]; we therefore

193 excluded pseudogenes from subsequent analyses. Significantly associated KEGG pathways included
194 "Spliceosome" (p=1.03e-8) and "Ubiquitin-mediated proteolysis" (p=2.4e-10) (Table 2).
195
196 **Table 2. Top 5 genes that were most significantly associated with mtDNA-CN within the "Spliceosome"**
197 **and "Ubiquitin-mediated proteolysis" KEGG pathways.**

Gene	Effect size estimate	Standard error	P-value
Spliceosome genes			
TRA2A	0.11	0.01	2.99e-14
LSM6	0.11	0.02	3.75e-10
HNRNPA1L2	0.12	0.02	2.07e-08
SRSF8	0.10	0.02	2.24e-07
NCBP2	0.06	0.01	6.60e-07
Ubiquitin-mediated proteolysis genes			
UBE2B	0.12	0.02	1.20e-13
ELOC	0.08	0.01	2.02e-08
UBE2I	0.09	0.02	6.26e-08
CUL1	0.07	0.01	9.73e-08
UBE2K	0.07	0.01	1.32e-07

198 **Effect size estimates represent the change in gene expression, in standard deviation units, associated**
199 **with a 1 standard deviation increase in blood-derived mtDNA-CN.**
200
201 A number of transcription factor target sequences were also significantly enriched, including those for
202 ELK1 (p=8.58e-66), NRF1 (p=1.76e-35), GABPB (p=3.54e-21), YY1 (p=3.14e-19), and E4F1 (p=3.98e-15).
203 All of these transcription factors regulate genes that play a role in mitochondrial function [35–39]. Gene
204 expression levels of these transcription factors were all positively correlated with mtDNA-CN, with 5 out
205 of 6 nominally significant, and 3 remaining significant after Bonferroni correction (p<8.33e-3) (Table 3).
206
207 **Table 3. Transcription factors whose targets are enriched for association with blood-derived mtDNA-**
208 **CN are nearly all nominally significantly associated with blood-derived mtDNA-CN.**

Gene	Effect size estimate (gene expression)	Standard error (gene expression)	P-value (gene expression)	P-value (enriched target sequences)
NRF1	0.03	0.02	0.07	1.76e-35
YY1	0.07	0.01	1.78e-06	3.14e-19

Gene	Effect size estimate (gene expression)	Standard error (gene expression)	P-value (gene expression)	P-value (enriched target sequences)
GABPB2	0.09	0.02	1.51e-09	3.54e-21
GABPB1	0.03	0.01	0.048	3.54e-21
E4F1	0.05	0.01	2.01e-04	3.98e-15
ELK1	0.04	0.02	0.021	8.58e-66

Effect size estimates, standard errors, and p-values from a linear regression between transcription factor gene expression and blood-derived mtDNA-CN. Transcription factors shown are those whose targets were significantly enriched for association with blood-derived mtDNA-CN.

Many mitochondrially-related cellular component gene ontology (GO) terms were significant, including "Mitochondrion" (p=7.77e-23), "Mitochondrial part" (p=2.79e-15), and "Mitochondrion organization" (p=2.87e-14) (Fig. 2) [40]. Additional significantly associated GO terms included "ubiquitin ligase complex" (p=6.6e-18) and "spliceosomal complex" (p=4.46e-14), supporting the KEGG pathway findings. Genes with substantial evidence of mitochondrial localization, determined through integration of several genome-scale datasets, were obtained from MitoCarta2.0 and demonstrated significant enrichment (p=8.22e-21) [41].

Cross-tissue analysis reveals associations between gene expression in multiple tissues and blood-derived mtDNA-CN

mtDNA-CN measured in blood has been associated with a number of aging-related diseases including chronic kidney disease, heart failure, and diabetes [10,42,43]. Given that these diseases primarily manifest in non-blood tissues, we evaluated associations between blood-derived mtDNA-CN and gene expression measured from 47 additional tissues that had greater than 50 samples after filtering. Though blood-derived mtDNA-CN appears to be associated with gene expression in other tissues, we did not observe a significant association between blood-derived mtDNA-CN and scaled mtRNA gene expression in any tissue other than blood, and only 2 out of 47 tested tissues had nominally significant associations between tissue-specific scaled mtRNA expression and blood-derived mtDNA-CN (Uterus [p

= 0.004], Heart – Left Ventricle [p = 0.017]). (Supp. Table 2). However, mtRNA expression for 35/47 non-blood tissues was positively associated with blood mtDNA-CN, which is more than what would be expected by chance ($p<0.001$). This suggests that while our study may be underpowered to detect a significant association in individual tissues due to small sample sizes, mtDNA-CN measured in blood is broadly correlated with mtDNA-CN in other tissues.

We calculated genomic inflation factors for each tissue to quantify test statistic inflation. Genomic inflation factors were elevated across multiple non-blood tissues, suggesting that blood-derived mtDNA-CN was broadly associated with gene expression in other tissues (Fig. 3). To determine true signal from noise, we performed 100 two-stage permutations for each tissue and obtained a genomic inflation factor lambda cutoff of >1.20 representing a significant elevation of lambda (study-wide $p<0.05$). Using this cutoff, we identified 30 non-blood tissues with a global inflation of test statistics [Supp. Table 3]. Other than blood, the most strongly enriched tissue was the putamen region of the brain, with a lambda of 3.27. We note that the two cell lines, EBV transformed lymphocytes (lambda=0.84) and cultured fibroblasts (lambda=0.84), showed no global inflation of test statistics, suggesting that blood-derived mtDNA-CN loses its association with gene expression after the cell-culturing process.

To examine the similarity of associations of mtDNA-CN observed in blood with other tissues, we calculated Spearman's rank correlation coefficients between effect estimates for blood-derived mtDNA-CN on blood gene expression (β_{blood}) and effect estimates for blood-derived mtDNA-CN on gene expression in other tissues (β_{tissue}). All genes that passed a permutation cutoff for significance in blood (p=2.7e-6, 721 genes total) were included. To distinguish tissues with correlations more extreme than baseline, we calculated reference correlations between blood and other tissues for randomly selected sets of genes. 26 tissues had observed values that were more extreme than the random gene sets (Supp. Table 4). Of these 26 tissues, 20 were among the 30 tissues with significantly inflated lambdas.

To identify pathways associated with mtDNA-CN across multiple tissues, we performed gene set enrichment analysis in each of the 30 tissues with a significant genomic inflation factor. Multiple terms were significant in greater than one tissue (Table 4), including terms related to oxidative phosphorylation and mitochondria, suggesting that mtDNA-CN derived from blood can reflect mitochondrial function occurring in other tissues.

Table 4. Pathways, transcription factor targets, and GO terms significantly enriched in multiple tissues.

Pathway	Number of significant tissues
Transcription factors	
SCGGAAGY_ELK1_02	18
RCGCANGCGY_NRF1_Q6	12
GCCATNTTG_YY1_Q6	8
TGCGCANK_UNKNOWN	8
MGGAAGTG_GABP_B	7
GO terms	
GO_RNA_BINDING	16
GO_CATALYTIC_COMPLEX	14
GO_CELLULAR_MACROMOLECULE_LOCALIZATION	13
GO_INTRACELLULAR_TRANSPORT	13
GO_MACROMOLECULE_CATABOLIC_PROCESS	12
KEGG terms	
KEGG_RIBOSOME	8
KEGG_HUNTINGTONS_DISEASE	4
KEGG_OXIDATIVE_PHOSPHORYLATION	4
KEGG_ALZHEIMERS_DISEASE	3
KEGG_PARKINSONS_DISEASE	3
Mitochondrial terms	
GO_MITOCHONDRIAL_ENVELOPE	8
GO_MITOCHONDRIAL_PART	8
GO_MITOCHONDRION	7
GO_MITOCHONDRION_ORGANIZATION	4
GO_MITOCHONDRIAL_PROTEIN_COMPLEX	3

The top 5 terms for each category that were significantly enriched in multiple tissues are shown.

ELK1 transcription factor binding sites were significantly enriched in 17 of the 30 significant tissues, and were also significant in whole blood, suggesting that mtDNA-CN may regulate ELK1 or vice versa. We note that gene expression for ELK1 was nominally significantly associated ($p<0.05$) with blood-derived

mtDNA-CN in 12 of the 18 tissues for which ELK1 targets were significantly enriched (Supp. Fig. 6). Effect estimates for ELK1 targets were generally consistent with the directionality of ELK1 effect estimates. For example, in blood, where ELK1 expression is positively associated with mtDNA-CN, 747/750 (99.6%) nominally significant ELK1 target genes were positively associated. On the other hand, mtDNA-CN was negatively associated with nerve ELK1 gene expression, and 204/306 (66.67%) nominally significant ELK1 target genes were also negatively associated. Of note, nearly all the noted transcription factors were ubiquitously expressed throughout the body, except for ELK1, which is not expressed in Brain Putamen or Spinal Cord (Supp. Fig. 7).

To identify genes driving enrichment of significant pathways in multiple tissues, we performed a random effects meta-analysis for all expressed genes using effect size estimates from all 47 non-blood tissues. Strikingly, genes encoding both the large and small ribosomal subunits were negatively associated with blood-derived mtDNA-CN across all tested tissues, implying an inverse relationship between ribosomal abundance and mitochondrial DNA quantity (Table 5).

Table 5. Random-effects meta-analysis for genes driving the enrichment of pathways in multiple tissues.

Gene	Meta effect size estimate	Meta standard error	Meta p-value
ELK1 targets			
STARD3	0.05	0.00	1.49e-17
EIF5A	0.08	0.01	4.97e-17
ERH	0.07	0.01	8.82e-17
RNA-binding genes			
SUZ12	0.06	0.00	4.49e-18
C1D	0.11	0.01	8.03e-18
MRPL23	-0.09	0.01	8.37e-18
Ribosome genes			
RPL34	-0.07	0.01	8.57e-16
RPS27	-0.08	0.01	1.35e-15
RPL39	-0.08	0.01	1.16e-14
Mitochondrial part genes			
MRPL23	-0.09	0.01	8.37e-18

Gene	Meta effect size estimate	Meta standard error	Meta p-value
MTERF3	-0.06	0.00	3.12e-17
MICU3	0.09	0.01	3.34e-17

Meta-analysis results are from all 47 tested tissues, excluding effects from whole blood. Top 3 most significant genes for each pathway are shown.

Huntington's disease (HD), Parkinson's disease (PD), and Alzheimer's disease (AD) were among the most significantly associated KEGG pathways that appear in multiple tissues (Table 5). This is an intriguing finding, given the known role of mitochondria in neurodegenerative disease [44].

While neurodegenerative disorders primarily manifest in nervous tissues [45], we observed significant enrichment of disease pathways in colon, pancreas, and testis tissues. When limiting our query to brain tissues, HD and PD were nominally significantly enriched in cerebellum, caudate (basal ganglia), and cortex, while AD was nominally significantly enriched in cerebellum and spinal cord (Supp. Fig 8).

mtDNA-CN is associated with incident neurodegenerative disease in the UKBiobank

To examine the association between mtDNA-CN and neurodegenerative disease risk, we used the UK Biobank (UKB) [46], a prospective cohort study with whole exome sequencing for ~50,000 individuals. mtDNA-CN was derived from whole exome sequencing and adjusted for sequencing artifacts, age, and sex. Analysis was restricted to individuals of European descent, and individuals with blood cell type count outliers were excluded. Using a Cox proportional-hazards model and adjusting for age and sex, we evaluated the relative risk of neurodegenerative disease associated with mtDNA-CN. Median follow-up time was approximately 10 years. Although the number of incident events was small, mtDNA-CN was significantly associated with Parkinson's disease (HR=0.75, CI=0.60;0.99) and Alzheimer's disease (HR=0.59, CI=0.44;0.81), and approached significance for non-Alzheimer's dementia (HR=0.81, CI=0.65;1.02). Consistent with other aging-related diseases [7,9], higher mtDNA-CN was associated with lower risk for developing incident neurodegenerative disease (Table 6). A combined analysis for all

304 individuals with incident neurodegenerative disease revealed a consistent strongly significant
305 association (HR=0.73, CI=0.66;0.90).
306
307 **Table 6. mtDNA-CN is associated with incident neurodegenerative disease.**

Disease	Hazard ratio	Confidence interval	Number of cases/controls	P-value
Parkinson's disease	0.75	0.60;0.99	63/39,044	0.030
Alzheimer's disease	0.59	0.44;0.81	41/39,100	0.001
Dementia (excluding AD)	0.81	0.65,1.02	74/39,048	0.074
Combined neurodegenerative disease	0.73	0.66;0.90	161/39,030	0.001

308 **Hazard ratios for incident neurodegenerative disease associate with a 1 standard deviation increase in**
309 **whole blood mtDNA-CN estimated from Cox proportional-hazards models in the UKBiobank. Analysis**
310 **was restricted to individuals of European descent, and individuals who were outliers for cell counts**
311 **were excluded from analysis.**
312
313 **DISCUSSION**
314 In this study, blood-derived mtDNA-CN was significantly associated with a host of blood-expressed
315 genes. As expected, nearly all genes involved in mtDNA replication were significantly associated with
316 mtDNA-CN in a positive direction. There was also a clear overall shift towards significant positive
317 estimates, possibly indicating that increased mtDNA-CN is reflective of a more active transcriptional
318 state. This finding is consistent with previous literature demonstrating that higher mitochondrial content
319 is correlated with increased transcriptional activity [47,48]. Strikingly, the two negatively associated
320 genes both play roles in innate immune function [41,42], suggesting that higher mtDNA-CN levels are
321 correlated with decreased immune response. Mitochondria play a role in immune responses to
322 pathogens in several ways; for example, mitochondrial DNA release from compromised mitochondria
323 can trigger an intracellular antiviral response through the cGAS/STING pathway [51], binding of viral
324 dsRNA to the mitochondrial antiviral signaling complex (MAVS) can trigger an interferon responses
325 through STAT6 activation [52], and release of mitochondrial components from cells can bind to damage-

associated molecular pattern (DAMP) receptors to trigger innate immune responses [53]. These novel findings correlating expression of mtDNA-CN with specific immune response genes in tissues represent an area for further investigation.

Gene set enrichment analyses revealed pathways potentially involved in mitochondrial DNA control, including ubiquitin-mediated proteolysis and splicing. Supporting this finding, Guantes et. al demonstrated that mitochondrial content modulates alternative splicing [47]. Additionally, we found that genes expressed in whole blood that were associated with blood-derived mtDNA-CN were enriched for target sequences for the ELK1, NRF1, YY1, GABPB, and E4F1 transcription factors. All of these transcription factors have been implicated in mitochondrial pathways, as ELK1 is associated with the mitochondrial permeability transition pore complex in neurons, NRF1 regulates expression of the mitochondrial translocase TOMM34, YY1 binds to and represses mitochondrial gene expression in skeletal muscle, GABPB is required for mitochondrial biogenesis, and E4F1 controls mitochondrial homeostasis [35–39]. Additionally, we found significant enrichment of signal for genes implicated in ubiquitin-mediated proteolysis and splicing. Given that mitochondrial quality control is regulated through ubiquitination, and that nuclear-encoded spliceosomes are involved in mtRNA splicing, our results likely implicate processes involved in mitochondrial DNA regulatory networks [54,55].

mtDNA-CN measured in one tissue has previously been found to be uncorrelated with mtDNA-CN in another tissue from the same individual [56]. We found that while mtRNA transcription in individual tissues was not significantly correlated with blood-derived mtDNA-CN, across all tissues, there was a significant enrichment for positive associations, suggesting a weak positive correlation between blood-derived mtDNA-CN and mtDNA-CN in other tissues. Moreover, we found that blood-derived mtDNA-CN was associated with various biological pathways in non-blood tissues (including mitochondrial function), providing a possible explanation as to why blood-derived mtDNA-CN is associated with aging-related diseases that primarily manifest in non-blood tissues. Further examination of pathways significant in

multiple tissues revealed that ribosomal subunit genes were significantly negatively associated with mtDNA-CN. While there has been conflicting evidence on the relationship between mtDNA-CN and ribosomal content, our study revealed a strong inverse relationship between ribosomal DNA dosage and mtDNA-CN [16,47]. Importantly, since these are statistical associations, causal directionality cannot be determined between gene expression and blood-derived mtDNA-CN. Future follow-up studies are needed to determine functional causality for mtDNA-CN and gene expression.

Strikingly, KEGG pathways that were significantly enriched in multiple tissues included Huntington's disease, Alzheimer's disease, and Parkinson's disease. These aging-related neurodegenerative diseases have all underlying mitochondrial pathologies [57–60] and dysregulated ubiquitination pathways [61]. In particular, mtDNA-CN has been implicated in Alzheimer's disease [62–64] and cognitive function [65,66]. Further, the ELK1 transcription factor, whose target sequences were significantly enriched in 18 tissues, plays a role in multiple neurodegenerative diseases [67]. Finally, after finding that blood-derived mtDNA-CN was associated with expression of neurodegenerative disease genes, we used an independent dataset, the UK Biobank, and found that mtDNA-CN was significantly associated with incident neurodegenerative disease risk. Notably, the population of individuals used for the UK Biobank analyses is biased towards fewer smokers and fewer prevalent Parkinson's events [68], which may impact the generalizability of our results. Despite this caveat, we show that blood-derived mtDNA-CN is significantly associated with gene expression from tissues across the body, and that higher mtDNA-CN is associated with decreased incident neurodegenerative disease risk.

METHODS

GTEx Sample acquisition

Whole genome sequences were downloaded from the GTEx version 8 cloud repository on 11/18/2020. RNA-sequencing data used for analyses was downloaded from the GTEx portal

(http://gtexportal.org/home/datasets) on 06/18/2019 and phenotypes were obtained from dbGaP (phs000424.**v8**.p2).

Estimation of mtDNA-CN

Samtools version 1.9 [69] was used to count the number of mitochondrial, unaligned, and total reads for each whole genome sequence. mtDNA-CN was estimated as the number of mitochondrial reads divided by the difference between the number of total reads and the number of unaligned reads to obtain a ratio of mtDNA to nuclear DNA. Whole genome is a highly accurate method for estimation of mtDNA-CN [22,56].

Correcting mtDNA-CN for covariates

All statistical analyses were performed with R version 3.6.1. Cell type composition for whole blood samples was determined from RNA-sequencing using xCell [19], only allowing for deconvolution of cell types found in blood. A stepwise regression in both directions was used to select appropriate cell types to correct mtDNA-CN. To avoid model overfitting, correlated cell types (R>0.8) were removed. The final model used to adjust mtDNA-CN included neutrophils, hematopoietic stem cells, megakaryocytes, subject cohort, ischemic time, age, and sex. Power calculations were performed using R^2 values from previous studies using the pwr package [22].

Filtering pipeline

A batch effect due to sample collection and/or sequencing methods resulted in significantly altered mtDNA-CN for individuals who were sequenced prior to January 2013. To keep this from confounding the analysis, we excluded subjects with whole genome sequencing prior to January 2013 (Supp. Fig. 1). Individuals who had greater than $5x10^7$ unaligned whole genome sequence reads were also omitted from the analysis. Cell type outliers who were greater than 3 standard deviations (SDs) from the mean were excluded as well. Only one individual remained from the surgical cohort after filtering and therefore was also removed (Supp. Fig. 9).

RNA-sequencing pipeline

GTEx version 8 RNA-sequencing data was downloaded from the GTEx website in transcripts per million (TPMs) and normalized using the trimmed mean of M-values method prior to analyses [70,71]. Genes with expression greater than 0.1 TPMs for at least 20% of samples were retained for analysis. To identify potential hidden confounders, we used surrogate variable analysis (SVA), protecting mtDNA-CN from SV generation [25]. SVs were associated with known covariates in the data, such as whether individuals were in the postmortem or the organ donor cohorts (Supp. Fig. 2). Individuals who were greater than 3 standard deviations from the mean for the first ten SVs were omitted from analysis. SV generation was performed iteratively 3 times.

Linear model for evaluating associations

To reduce the influence of outliers, both the gene expression metric and the mtDNA-CN metric were inverse normal transformed prior to linear regression. We then tested for association using multiple linear regression, with mtDNA-CN as the predictor and gene expression as the outcome, correcting for SVs, sex, cohort, race, ischemic time, and the first three genotyping principal components.

Genomic inflation factor calculation

Genomic inflation factors were calculated by squaring z-scores to obtain chi-squared values. The median observed chi-squared value was divided by the expected median to obtain lambda [26].

Two-stage permutations

To determine an appropriate p-value cutoff, we created null datasets for permutation testing. First, a multiple linear regression model for the alternate hypothesis was used to obtain gene expression residuals. Second, a multiple linear regression model for the null hypothesis was used to obtain estimates for each gene. Residuals from the alternate model were than permuted and added to effect estimates from the null model to create null datasets. Permuted gene expression data was then tested for association with mtDNA-CN. Unless otherwise stated, permutations were performed 100 times.

Minimum p-values from each permuted dataset were obtained, and the 5th lowest p-value was utilized as a permutation cutoff.

Annotation of gene categories

Gene annotations were downloaded from Gencode [27]. Test statistics were then stratified by gene type and observed and expected distributions were generated for each category.

Overrepresentation of positive beta estimates

Percentage of positive effect estimates was calculated using all nominally significant genes in blood, dividing the number of nominally significant genes with positive effect estimates by the total number of nominally significant genes. Percentages for null distributions were calculated using 1000 permutations generated using the two-stage permutation method described above.

Gene set enrichment analysis

To examine enrichment for genes in specific pathways, gene sets for KEGG pathways, transcription factor target sequences, and gene ontologies were downloaded from the Molecular Signatures database [25,27,49]. Then, using the absolute value of the t-scores from the regression model with mtDNA-CN, we performed a t-test of t-scores for genes in a specific pathway versus genes that were not contained in the pathway. Permutations using randomized t-scores were used to determine appropriate cutoffs for significance. To confirm that results were not driven by individual genes in a pathway with very large t-scores, we also performed t-tests using ranked t-scores as opposed to absolute value t-scores.

REVIGO trimming and visualization of GO terms

For visualization of significantly enriched GO terms and elimination of redundant GO terms, REVIGO (http://revigo.irb.hr/) was used with the default settings except for the allowed similarity, which was set to medium (0.7) [40].

Testing for associations between blood-derived mtDNA-CN and gene expression in other tissues

Filtering parameters and models for testing the association of blood-derived mtDNA-CN with gene expression in other tissues were identical to the pipeline used in whole blood. Only tissues with greater than 50 observations after filtering were tested. For tissues that had no variation in covariates, covariates were dropped from the linear model (i.e. sex was not used in the model for testing gene expression in reproductive organs and cohort was not used in the model for brain tissues).

Spearman correlations for effect estimates with whole blood

All significant genes in whole blood that passed the permutation cutoff ($p=2.7x10^{-6}$) were used for testing. Spearman correlations between effect estimates in blood and effect estimates in other tissues were calculated. To compare correlations for genes significant in blood with baseline correlation, we randomly selected 100 random genes and calculated correlations between blood estimates and specific tissue estimates for those genes. We repeated this random selection 100 times to generate multiple baseline correlation measures.

Meta-analysis of genes driving specific ontologies

To calculate meta-analysis effect estimates and p-values, the R '*meta*' package [73] was used to perform a random-effects meta-analysis using all effect estimates and p-values for all tissues, excluding results from whole blood.

Association of mtDNA-CN with neurodegenerative disease in UKB

Samtools version 1.9 was used to extract read summary statistics from 49,997 UK biobank whole exome sequences. An in-house perl script was then used to aggregate summary statistics into the number of total, mapped, unmapped, autosomal, chromosome X, chromosome Y, mitochondrial, random, unknown, decoy1, and decoy2 reads. 10-fold cross validation was used to select linear regression covariates to adjust the number of mitochondrial reads for. After correcting for these potential technical artifacts, this metric was then adjusted for age and sex. Cell type outliers were excluded from the dataset, and subsequent analyses were restricted to individuals of European descent. A Cox

proportional-hazards model was used to evaluate the association between mtDNA-CN and time to incident neurodegenerative disease, adjusting for age and sex.

DECLARATIONS

Ethics approval and consent to participate

Approval access for the datasets in this study was obtained from the GTEx and UKBiobank resources.

Consent for publication

Not applicable

Availability of data and materials

RNA-sequencing data used for analyses was downloaded from the GTEx portal (http://gtexportal.org/home/datasets) on 06/18/2019 and phenotypes were obtained from dbGaP (phs000424.**v8**.p2). Aligned whole genome sequences were downloaded from google cloud. UKBiobank data was accessed under application number 17731. All in-house scripts can be found in the following Github repository: htpps://github.com/syyang93/mtDNA_GE_scripts.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants R01HL13573 and R01HL144569.

Authors' contributions

Concept and design: Yang, Arking

Acquisition, analysis, or interpretation of data: Yang, Castellani, Longchamps, Pillalamarri, Arking

Drafting manuscript: Yang, Arking

Critical revision of the manuscript: Yang, Castellani, Longchamps, Pillalamarri, O'Rourke, Guallar, Arking

Obtaining funding: Arking

Acknowledgements

This research was conducted using data from the Genotype-Tissue Expression (GTEx) project (dbGaP accession: phs000424.**v8**.p2). The GTEx project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

This research was also conducted using the UK Biobank Resource under Application Number 17731.

REFERENCES

1. Wai T, Ao A, Zhang X, Cyr D, Dufort D, Shoubridge EA. The role of mitochondrial DNA copy number in mammalian fertility. Biol Reprod. 2010;83:52–62.

2. Clay Montier LL, Deng J, Bai Y. Number matters: control of mammalian mitochondrial DNA copy number. J Genet Genomics. 2009;36:125–31.

3. Miller FJ, Rosenfeldt FL, Zhang C, Linnane AW, Nagley P. Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. Nucleic Acids Res. 2003;31:e61.

4. Herst PM, Rowe MR, Carson GM, Berridge MV. Functional Mitochondria in Health and Disease. Front Endocrinol (Lausanne) [Internet]. 2017 [cited 2020 May 28];8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5675848/

5. Malik AN, Czajka A. Is mitochondrial DNA content a potential biomarker of mitochondrial dysfunction? Mitochondrion. 2013;13:481–92.

6. Castellani CA, Longchamps RJ, Sun J, Guallar E, Arking DE. Thinking outside the nucleus: Mitochondrial DNA copy number in health and disease. Mitochondrion. 2020;53:214–23.

7. Ashar FN, Moes A, Moore AZ, Grove ML, Chaves PHM, Coresh J, et al. Association of mitochondrial DNA levels with frailty and all-cause mortality. J Mol Med. 2015;93:177–86.

8. Association between Mitochondrial DNA Copy Number in Peripheral Blood and Incident CKD in the Atherosclerosis Risk in Communities Study | American Society of Nephrology [Internet]. [cited 2020 Feb 17]. Available from: https://jasn.asnjournals.org/content/27/8/2467

9. Ashar FN, Zhang Y, Longchamps RJ, Lane J, Moes A, Grove ML, et al. Association of Mitochondrial DNA Copy Number With Cardiovascular Disease. JAMA Cardiol. 2017;2:1247–55.

10. Huang J, Tan L, Shen R, Zhang L, Zuo H, Wang DW. Decreased Peripheral Mitochondrial DNA Copy Number is Associated with the Risk of Heart Failure and Long-term Outcomes. Medicine (Baltimore) [Internet]. 2016 [cited 2020 Feb 17];95. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4839823/

11. Ali AT, Boehme L, Carbajosa G, Seitan VC, Small KS, Hodgkinson A. Nuclear genetic regulation of the human mitochondrial transcriptome. McCarthy MI, Battle A, Arking D, editors. eLife. 2019;8:e41927.

12. Lee WT, Sun X, Tsai T-S, Johnson JL, Gould JA, Garama DJ, et al. Mitochondrial DNA haplotypes induce differential patterns of DNA methylation that result in differential chromosomal gene expression patterns. Cell Death Discov. 2017;3:17062.

13. Sun X, St John JC. Modulation of mitochondrial DNA copy number in a model of glioblastoma induces changes to DNA methylation and gene expression of the nuclear genome in tumours. Epigenetics Chromatin. 2018;11:53.

14. Reznik E, Miller ML, Şenbabaoğlu Y, Riaz N, Sarungbam J, Tickoo SK, et al. Mitochondrial DNA copy number variation across human cancers. Elife. 2016;5.

15. Castellani CA, Longchamps RJ, Sumpter JA, Newcomb CE, Lane JA, Grove ML, et al. Mitochondrial DNA Copy Number (mtDNA-CN) Can Influence Mortality and Cardiovascular Disease via Methylation of Nuclear DNA CpGs [Internet]. Genomics; 2019 Jun. Available from: http://biorxiv.org/lookup/doi/10.1101/673293

16. Gibbons JG, Branco AT, Yu S, Lemos B. Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans. Nature Communications. 2014;5:1–12.

17. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580–5.

18. Zhang R, Wang Y, Ye K, Picard M, Gu Z. Independent impacts of aging on mitochondrial DNA quantity and quality in humans. BMC Genomics [Internet]. 2017 [cited 2020 Mar 4];18. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5697406/

19. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017;18:220.

20. Moore AZ, Ding J, Tuke MA, Wood AR, Bandinelli S, Frayling TM, et al. Influence of cell distribution and diabetes status on the association between mitochondrial DNA copy number and aging phenotypes in the InCHIANTI study. Aging Cell [Internet]. 2018 [cited 2020 Feb 17];17. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5770782/

21. Mengel-From J, Thinggaard M, Dalgård C, Kyvik KO, Christensen K, Christiansen L. Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly. Hum Genet. 2014;133:1149–59.

22. Longchamps RJ, Castellani CA, Yang SY, Newcomb CE, Sumpter JA, Lane J, et al. Evaluation of mitochondrial DNA copy number estimation techniques. PLOS ONE. 2020;15:e0228166.

23. Garcia I, Jones E, Ramos M, Innis-Whitehouse W, Gilkerson R. The little big genome: the organization of mitochondrial DNA. Front Biosci (Landmark Ed). 2017;22:710–21.

24. Rusecka J, Kaliszewska M, Bartnik E, Tońska K. Nuclear genes involved in mitochondrial diseases caused by instability of mitochondrial DNA. J Appl Genet. 2018;59:43–57.

25. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 2007;3:1724–35.

563 26. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55:997–1004.

564 27. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: The
565 reference human genome annotation for The ENCODE Project. Genome Res. 2012;22:1760–74.

566 28. Irizarry RA, Wang C, Zhou Y, Speed TP. Gene Set Enrichment Analysis Made Simple. Stat Methods
567 Med Res. 2009;18:565–75.

568 29. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular
569 signatures database (MSigDB) 3.0. Bioinformatics. Oxford Academic; 2011;27:1739–40.

570 30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the
571 unification of biology. Nat Genet. 2000;25:25–9.

572 31. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, et al. Systematic discovery of regulatory
573 motifs in human promoters and 3' UTRs by comparison of several mammals. Nature. Nature Publishing
574 Group; 2005;434:338–45.

575 32. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. Oxford Academic;
576 2019;47:D330–8.

577 33. Meng H, Yaari G, Bolen CR, Avey S, Kleinstein SH. Gene set meta-analysis with Quantitative Set
578 Analysis for Gene Expression (QuSAGE). PLOS Computational Biology. Public Library of Science;
579 2019;15:e1006899.

580 34. Saha A, Battle A. False positives in trans-eQTL and co-expression analyses arising from RNA-
581 sequencing alignment errors. F1000Res [Internet]. 2019 [cited 2020 Mar 4];7. Available from:
582 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6305209/

583 35. Chen F, Zhou J, Li Y, Zhao Y, Yuan J, Cao Y, et al. YY1 regulates skeletal muscle regeneration through
584 controlling metabolic reprogramming of satellite cells. EMBO J. 2019;38.

585 36. Barrett LE, Bockstaele EJV, Sul JY, Takano H, Haydon PG, Eberwine JH. Elk-1 associates with the
586 mitochondrial permeability transition pore complex in neurons. PNAS. National Academy of Sciences;
587 2006;103:5155–60.

588 37. Blesa JR, Prieto-Ruiz JA, Abraham BA, Harrison BL, Hegde AA, Hernández-Yago J. NRF-1 is the major
589 transcription factor regulating the expression of the human TOMM34 gene. Biochem Cell Biol.
590 2008;86:46–56.

591 38. Yang Z-F, Drumea K, Mott S, Wang J, Rosmarin AG. GABP transcription factor (nuclear respiratory
592 factor 2) is required for mitochondrial biogenesis. Mol Cell Biol. 2014;34:3194–201.

593 39. Rodier G, Kirsh O, Baraibar M, Houlès T, Lacroix M, Delpech H, et al. The Transcription Factor E4F1
594 Coordinates CHK1-Dependent Checkpoint and Mitochondrial Functions. Cell Reports. 2015;11:220–33.

595 40. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene
596 Ontology Terms. PLOS ONE. Public Library of Science; 2011;6:e21800.

41. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins [Internet]. [cited 2020 Feb 25]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4702768/

42. Al-Kafaji G, Aljadaan A, Kamal A, Bakhiet M. Peripheral blood mitochondrial DNA copy number as a novel potential biomarker for diabetic nephropathy in type 2 diabetes patients. Exp Ther Med. 2018;16:1483–92.

43. Tin A, Grams ME, Ashar FN, Lane JA, Rosenberg AZ, Grove ML, et al. Association between Mitochondrial DNA Copy Number in Peripheral Blood and Incident CKD in the Atherosclerosis Risk in Communities Study. J Am Soc Nephrol. 2016;27:2467–73.

44. Reddy PH. The role of mitochondria in neurodegenerative diseases: mitochondria as a therapeutic target in Alzheimer's disease. CNS Spectr. 2009;14:8–18.

45. Wood LB, Winslow AR, Strasser SD. Systems Biology of Neurodegenerative Diseases. Integr Biol (Camb). 2015;7:758–75.

46. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. Nature Publishing Group; 2018;562:203–9.

47. Guantes R, Rastrojo A, Neves R, Lima A, Aguado B, Iborra FJ. Global variability in gene expression and alternative splicing is modulated by mitochondrial content. Genome Res. 2015;25:633–44.

48. Márquez-Jurado S, Díaz-Colunga J, das Neves RP, Martinez-Lorente A, Almazán F, Guantes R, et al. Mitochondrial levels determine variability in cell death by modulating apoptotic gene expression. Nature Communications. Nature Publishing Group; 2018;9:389.

49. Osanai A, Sashinami H, Asano K, Li S-J, Hu D-L, Nakane A. Mouse Peptidoglycan Recognition Protein PGLYRP-1 Plays a Role in the Host Innate Immune Response against Listeria monocytogenes Infection. Infection and Immunity. American Society for Microbiology Journals; 2011;79:858–66.

50. Gombart AF, Borregaard N, Koeffler HP. Human cathelicidin antimicrobial peptide (CAMP) gene is a direct target of the vitamin D receptor and is strongly up-regulated in myeloid cells by 1,25-dihydroxyvitamin D3. FASEB J. 2005;19:1067–77.

51. West AP, Khoury-Hanold W, Staron M, Tal MC, Pineda CM, Lang SM, et al. Mitochondrial DNA stress primes the antiviral innate immune response. Nature. 2015;520:553–7.

52. Chen H, Sun H, You F, Sun W, Zhou X, Chen L, et al. Activation of STAT6 by STING is critical for antiviral innate immunity. Cell. 2011;147:436–46.

53. Nakahira K, Hisata S, Choi AMK. The Roles of Mitochondrial Damage-Associated Molecular Patterns in Diseases. Antioxid Redox Signal. 2015;23:1329–50.

54. Bragoszewski P, Turek M, Chacinska A. Control of mitochondrial biogenesis and function by the ubiquitin–proteasome system. Open Biol [Internet]. 2017 [cited 2020 Mar 31];7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5413908/

631 55. Herai RH, Negraes PD, Muotri AR. Evidence of nuclei-encoded spliceosome mediating splicing of
632 mitochondrial RNA. Hum Mol Genet. 2017;26:2472–9.

633 56. Wachsmuth M, Hübner A, Li M, Madea B, Stoneking M. Age-Related and Heteroplasmy-Related
634 Variation in Human mtDNA Copy Number. PLOS Genetics. 2016;12:e1005939.

635 57. Petersen MH, Budtz-Jørgensen E, Sørensen SA, Nielsen JE, Hjermind LE, Vinther-Jensen T, et al.
636 Reduction in mitochondrial DNA copy number in peripheral leukocytes after onset of Huntington's
637 disease. Mitochondrion. 2014;17:14–21.

638 58. Coskun PE, Wyrembak J, Derbereva O, Melkonian G, Doran E, Lott IT, et al. Systemic mitochondrial
639 dysfunction and the etiology of Alzheimer's disease and down syndrome dementia. J Alzheimers Dis.
640 2010;20 Suppl 2:S293-310.

641 59. Park J-S, Davis RL, Sue CM. Mitochondrial Dysfunction in Parkinson's Disease: New Mechanistic
642 Insights and Therapeutic Perspectives. Curr Neurol Neurosci Rep [Internet]. 2018 [cited 2020 May
643 19];18. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5882770/

644 60. Wei W, Keogh MJ, Wilson I, Coxhead J, Ryan S, Rollinson S, et al. Mitochondrial DNA point mutations
645 and relative copy number in 1363 disease and control human brains. Acta Neuropathol Commun
646 [Internet]. 2017 [cited 2020 Jun 8];5. Available from:
647 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5290662/

648 61. Atkin G, Paulson H. Ubiquitin pathways in neurodegenerative disease. Front Mol Neurosci [Internet].
649 2014 [cited 2020 May 19];7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4085722/

650 62. Rice AC, Keeney PM, Algarzae NK, Ladd AC, Thomas RR, Bennett JP. Mitochondrial DNA copy
651 numbers in pyramidal neurons are decreased and mitochondrial biogenesis transcriptome signaling is
652 disrupted in Alzheimer's disease hippocampi. J Alzheimers Dis. 2014;40:319–30.

653 63. Delbarba A, Abate G, Prandelli C, Marziano M, Buizza L, Arce Varas N, et al. Mitochondrial Alterations
654 in Peripheral Mononuclear Blood Cells from Alzheimer's Disease and Mild Cognitive Impairment
655 Patients. Oxid Med Cell Longev. 2016;2016:5923938.

656 64. Lv X, Zhou D, Ge B, Chen H, Du Y, Liu S, et al. Association of Folate Metabolites and Mitochondrial
657 Function in Peripheral Blood Cells in Alzheimer's Disease: A Matched Case-Control Study. J Alzheimers
658 Dis. 2019;70:1133–42.

659 65. Lee J-W, Park KD, Im J-A, Kim MY, Lee D-C. Mitochondrial DNA copy number in peripheral blood is
660 associated with cognitive function in apparently healthy elderly women. Clin Chim Acta. 2010;411:592–
661 6.

662 66. Lee J-Y, Kim J-H, Lee D-C. Combined Impact of Telomere Length and Mitochondrial DNA Copy
663 Number on Cognitive Function in Community-Dwelling Very Old Adults. DEM. Karger Publishers;
664 2017;44:232–43.

665 67. Besnard A, Galan-Rodriguez B, Vanhoutte P, Caboche J. Elk-1 a Transcription Factor with Multiple
666 Facets in the Brain. Front Neurosci [Internet]. 2011 [cited 2020 Jun 19];5. Available from:
667 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3060702/

668 68. Van Hout CV, Tachmazidou I, Backman JD, Hoffman JX, Ye B, Pandey AK, et al. Whole exome
669 sequencing and characterization of coding variation in 49,960 individuals in the UK Biobank [Internet].
670 Genomics; 2019 Mar. Available from: http://biorxiv.org/lookup/doi/10.1101/572347

671 69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map
672 format and SAMtools. Bioinformatics. 2009;25:2078–9.

673 70. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-
674 seq data. Genome Biology. 2010;11:R25.

675 71. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression
676 analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.

677 72. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res.
678 2000;28:27–30.

679 73. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid
680 Based Mental Health. 2019;22:153–60.

681

Figure 1. Global inflation of test statistics from linear regressions between blood-derived mtDNA-CN and gene expression in blood. After stratification by gene category, protein-coding genes have the most inflation, suggesting that mtDNA-CN is strongly associated with genes that code for proteins.

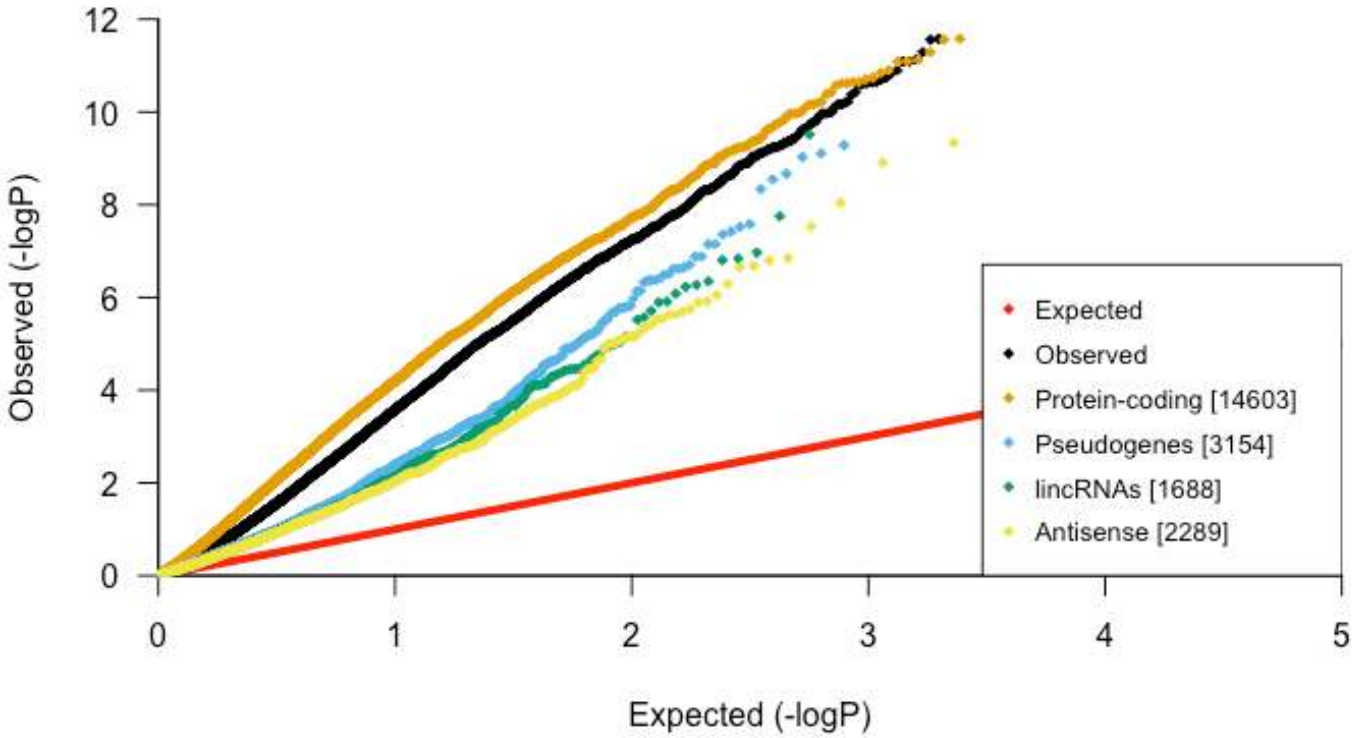

Figure 2. REVIGO visualization of GO Cellular Component terms significantly associated with mtDNA-CN after removal of redundant GO terms. Size of the circle represents the relative number of genes in each gene set, color represents significance. Axis represent semantic similarities between GO terms; GO terms that are more similar will cluster with one another.

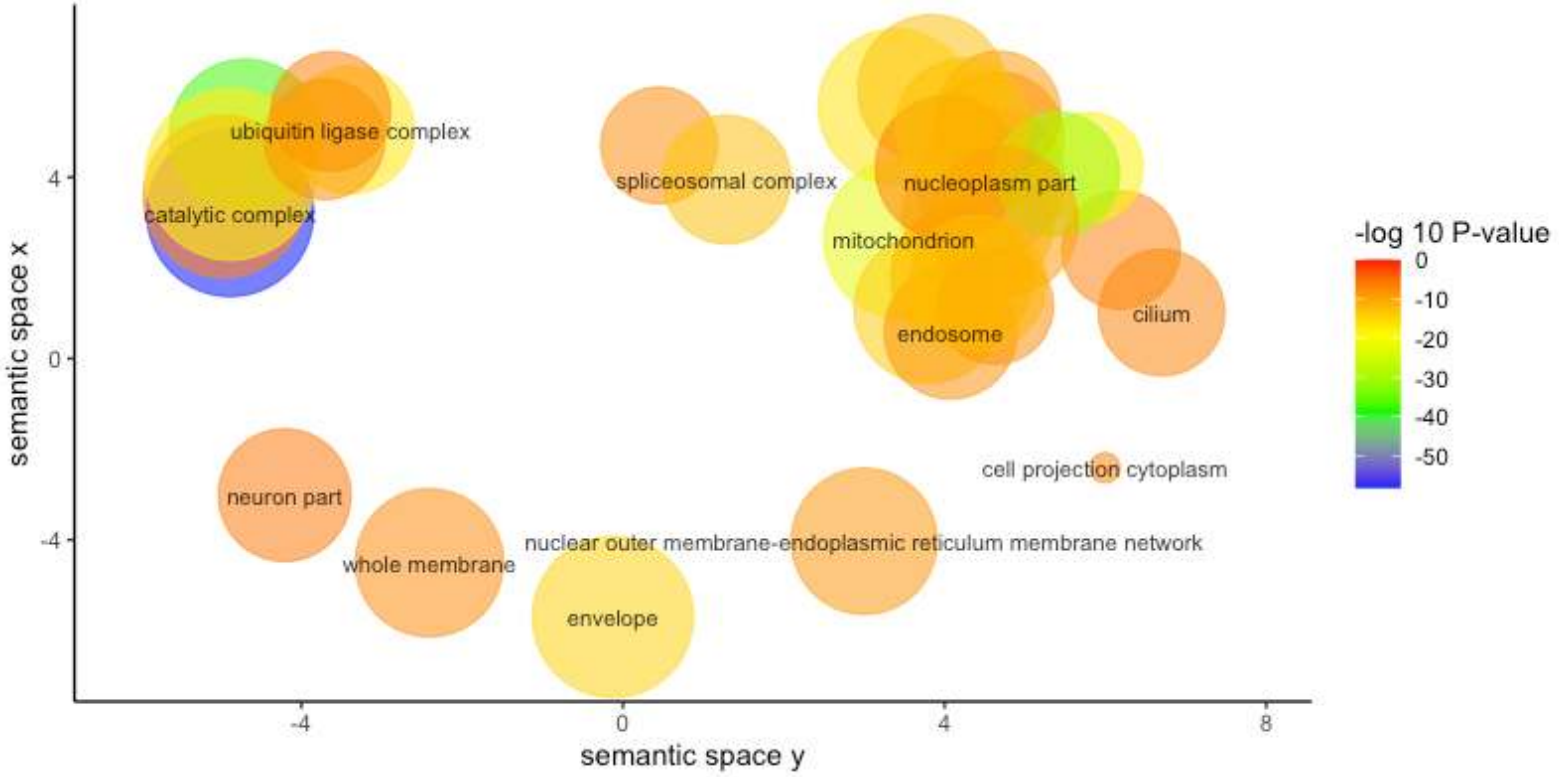

Figure 3. Observed genomic inflation factors are significantly different from permuted genomic inflation factors for certain tissues. Higher genomic inflation factor represents increased global associations between blood-derived mtDNA-CN and gene expression in a specific tissue. Permuted genomic inflation factors were obtained using two-stage permutation testing.